Automated Data Processing (ADP) Research and Development

Farid U. Dowla, Virgil N. Kohlhepp, and Richard R. Leach, Jr.

Earth Sciences Division, Lawrence Livermore National Laboratory

Sponsored by DOE CTBT R&D Program[1]

Abstract

Monitoring a comprehensive test ban treaty (CTBT) will require screening tens of thousands of seismic events each year. Reliable automated data analysis will be essential in keeping up with the continuous stream of events that a global monitoring network will detect. We are developing automated event location and identification algorithms by looking at the gaps and weaknesses in conventional ADP systems and by taking advantage of modern computational paradigms. Our research focus is on three areas: developing robust algorithms for signal feature extraction, integrating the analysis of critical measurements, and exploiting joint estimation techniques such as using data from acoustic, hydroacoustic, and seismic sensors. We identify several important problems for research and development; e.g., event location with approximate velocity models and event identification in the presence of outliers. We are employing both linear and nonlinear methods and advanced signal transform techniques to solve these event monitoring problems. Our goal is to increase event-interpretation throughput by employing the power and efficiency of modern computational techniques, and to improve the reliability of automated analysis by reducing the rates of false alarms and missed detections.

Key words: Automation, data fusion, event location, event identification, neural networks, self-organization, seismic networks, seismic arrays, signal processing, wavelets.

[1] Research performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract W-7405-ENG-48.

1. OBJECTIVES

Our primary objective is to develop efficient and reliable automated event location and identification algorithms. We are currently focusing on the problem of locating and identifying low-magnitude events. Only a small number of regional stations detect these events, but we must screen tens of thousands of such events each year. We emphasize three areas of research and development in the automatic processing of this data. First, how can we extract event features from full waveforms? Second, what is the best way to organize and integrate critical signal measurements? Finally, what improvements to seismic event detection can be realized by synergistic use of data from diverse sensors (e.g., seismic, hydroacoustic, infrasound, and satellite)? We must make progress on all these fronts to significantly improve the performance of the current monitoring systems in terms reduced rates of false alarms and missed detections.

2. PRELIMINARY RESEARCH RESULTS

2.1. Extraction of Features from Full Waveforms

2.1.1 Wavelet Analysis. The wavelet transform $W(a,\tau)$ of a signal, $s(t)$, such as a seismogram, is given by

$$W(a,\tau)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} s(t)\ h(\ [t-\tau]/\ a\ )\,dt,$$

where a represents the scale and τ time in the transform domain, and $h(t)$ is the analysis wavelet, which is localized in both time and frequency (Foufoula-Georgio and Kumar, 1994). It is well known that the most important features in a pattern are those that persist through different scales. Although computing the wavelet transform is somewhat similar to bandpass filtering, an appropriately designed wavelet brings out the concept of scale explicitly and it is particularly useful for transient analysis because of its constant-Q property. We have designed wavelets starting from real seismograms; Figure 1 shows example wavelet coefficients (in the transform domain) and bandpass filtering versions of a seismogram as a function of scale and time. The wavelet transform is an efficient representation of the seismogram and preserves the significant features in the waveform. We are developing noise-cancellation beamforming algorithms making use of this transform. The wavelet transform is an invertible transform meaning that the signal can be recovered perfectly from the wavelet coefficients. Note that unlike Fourier coefficients, wavelet coefficients preserve shape in the transform domain. Furthermore, while bandpass filtering represents each band of a N-point signal with another N-point signal, the discrete wavelet transform represents all scales with just one N-point set. Hence, the wavelet transform is more efficient than its corresponding filter bank representation.

2.1.2 High-Resolution Model Based Spectrograms. A spectrogram is simply a series of spectra computed by segmenting a long time series into many small and overlapping segments. We find that when the spectral estimation uses a high-resolution autoregressive model, spectrograms can play an important role in extracting the time-frequency features of the signal. The rapidly varying local features of the signal appear as functions of the frequency variable, and the slowly varying global trends of the signal appear as functions of the time variable. Finally, the large time-bandwidth product captured by a spectrogram helps in distinguishing the many different types of seismic events.

2.2 Organization and Integration of Signal Measurements

2.2.1 Event Identification with an Ill-Distributed Database. A rather stiff requirement on an ADP system is that it be able to detect and identify events from new source regions. This is a particularly difficult problem because we usually develop event identification systems using "recognition" or match-filtering techniques. In this section we discuss a solution to a slightly less difficult and related problem, where the number of explosions is much less than the number earthquakes in the training database, a realistic treaty-monitoring scenario.

Consider the problem of discriminating a nuclear explosion from an earthquake. Instead of attempting to solve a "binary" (i.e., earthquake=0, explosion=1) discrimination problem, suppose

we design an algorithm that can classify events by their various features; i.e., events are distinguished by depth, faulting mechanism, magnitude, and the like. We find that this type of n-ary discrimination algorithm is quite useful for outlier identification and might indeed be able to handle "new" or "unseen" event types because the projection of the "new" event ought to be poor on all the classes of events the method is able to recognize. A binary discrimination algorithm is forced to ignore differences that might exist among the events in a particular class. This results in its inability to distinguish outliers. Because there are many different types of earthquakes and explosions, we first employ self-organization of all the events in the database. This is followed by a match-filtering technique.

As shown in Figure 2-a, the identification system is a two-stage process. First, we allow all events in the database to self-organize into different cluster groups according to the similarity of the spectrograms. We use the self-organizing neural network to solve this clustering problem, although other clustering algorithms might work just as well. Members of some of these cluster groups will be purely earthquakes while other clusters might consist of just explosions; on the other hand, some clusters will consist of a mixture of earthquakes and explosions. In the second stage, we introduce a test event. If this test event has low correlation (poor match) with all cluster centroids or means, that event is an outlier. If the test event falls into one of the homogenous cluster groups with high correlation, there is no ambiguity about its class type. However, if a new event falls into a mixed class, we use only the members of that class and their identities to classify the test event. For example, the hierarchical clustering technique used by Israelsson (1991) might be quite appropriate at this stage because the number of events in any one cluster is reduced after the process of self-organization. We have completed a case study using earthquakes and explosion data from the NTS area. We used bootstrapping techniques and obtained results as function of the number of explosions in the training set. The results, shown in Figure 2-b, compare favorably against a standard Pg/Lg receiver operating curve.

2.2.2. Numerical Optimization and Global Search Techniques. The application of robust and fast numerical optimization techniques is essential with large volumes of high-dimensional data, a problem we face with seismic ADP. We developed optimization algorithms (Johansson et al., 1992; Altschuler et al., 1994) for nonlinear least-squares error minimization and for constrained global optimization techniques. We have successfully used these algorithms in many seismic event identification problems including discrimination and yield estimation (Dowla and Rogers, 1995). These methods will also be useful for event location applications, since location problems are solved by a global search method (Shearer, 1994).

2.3 Synergistic Use of Seismic, Acoustic, and Hydroacoustic Data

To improve our ability to verify adherence to treaties limiting or banning nuclear tests, we may need to combine seismic data with data from other types of sensors. Event identification that combines data from a diverse range of sensor types, such as seismic, hydroacoustic, infrasound, optical, or acoustic sensors, has been discussed recently. We are exploring the possibility of synergistic use of seismic, acoustic, and hydroacoustic data in event location and identification.

2.3.1 Acoustic and Seismic. One area of concern is the threat of clandestine nuclear testing under the guise of a normally operating mine. An inability to distinguish a nuclear test from a typical mining blast would constitute a serious problem in treaty monitoring. A potentially powerful approach to resolving these issues is to supplement seismic monitoring with acoustic analysis of mining blasts. Acoustic signals might be more noisy at low frequencies; for example, wind and storms often generate low frequency noise. However, for short distances, acoustic signals do not suffer from the strong attenuation of high frequency propagation of seismic signals. Combining both acoustic and seismic signals could improve the monitoring capability at a suspect mine. In addition, the cost of complementing seismic with acoustic monitoring is relatively low.

We are beginning to study seismic and acoustic signals recorded from two mining events that occurred in the Newmont Gold Company (NGC) gold quarry near Carlin, Nevada. A map of the quarry is shown in Figure 3. Example acoustic and seismic spectrograms are shown in Figures 4-a and 4-b.

Although seismic spectrogram analysis often works well in classifying mining events, there are instances for which this method becomes inadequate. One example is when ripple-fired, time-delayed detonations occur very close to each other (i.e., when mining operations use very short firing delays, on the order of 15 ms or so). These short delays result in higher frequencies at which spectral scalloping is observed. Because high-frequency signals are strongly attenuated, it becomes more and more difficult to identify the scalloping in the spectra of the recorded data. Depending on the geology, significant high frequency attenuation may occur even over short distances. In this situation, time-frequency features in the acoustic signal might be useful for source identification. In summary, ripple fire characteristics appear to be quite prominent in the spectrogram of the acoustic records and we think that the near source acoustic records are useful in the monitoring of mine shots.

2.3.2 Hydroacoustic and Seismic. Because wave propagation velocities are quite different for hydroacoustic and seismic signals, event location based on travel time analysis can be significantly more robust and accurate when travel time data are combined from the two types of signals. We are developing a database containing recordings from both hydroacoustic and seismic sensors for the same events. An example of a seismic event recorded using both types of sensors is shown in Figure 5. Since the source is the same for the two waveforms, similarities and differences between the two data can be useful in source characterization.

3. RECOMMENDATIONS AND FUTURE PLANS

Accurate event location for small magnitude events is a high priority problem for LLNL ADP research and development efforts. Some of the problems on which we are focusing are outlined in the following.

Given the fact that regional velocity models are not precisely known, we must develop location algorithms that are less sensitive to propagation models. While an *approximate* event hypocenter can be determined quite rapidly with a network of stations, *accurate* determination of a hypocenter is nontrivial (Herrmann, 1982; Veith, 1985; Palvis, 1986; Thurber, 1985; Wirth et al., 1976; Flinn 1965). Given arrivals at a number of stations and an approximate event location, the determination of the hypocenter estimate might be improved by judicial use of network travel time data and by extracting selected events from the database and performing station corrections based only on selected events and on the new event, thus minimizing uncertainties in event location. Fast global search algorithms must be developed both to select the events and to minimize arrival time residuals in determining event location. The feasibility of improving the velocity model while solving for the event location must be explored. Arrival time information from the secondary phases and bearing estimates from arrays and three-component stations should be incorporated into the algorithm. Finally, depth estimation with an optimal choice of stations might prove to be useful.

While location by waveform correlation has been proposed by many authors (Harris, 1991; Israelsson, 1990; Riviere-Barber, 1993; Thorbjarnardottir, 1987), there does not exist a robust algorithm that uses both waveform correlation and travel-time onset analysis and that converges to a consistent solution. Somewhat similar to the PMCC method (Cansi, 1995), it might be possible to systematically use stations of the network and to use advanced correlation techniques (Park et al., 1987; Duckworth, 1987) on full waveforms for improved location. We plan to explore the use of various error norms for outlier removal. In addressing the correlation estimation problem, the distance of the events from the stations makes the normalization issue an important one; i.e., normalized correlation (Neidell and Taner, 1971). Correlation as a function of frequency and scale (or wavelets) must to be studied.

Although a highly challenging problem, the location of events in an earthquake swarm or during mine related activities is important in CTBT monitoring. Very few studies have addressed this problem in depth in spite of the fact that treaty violations could easily go undetected if a test were performed in the midst of earthquake aftershocks or during quarry blasts and mine collapses. Joint use of bearing, amplitude, and travel time data might be necessary to develop location algorithms will exhibit robust performance.

Finally, use of reliable ground-truth data is important in developing better algorithms. For example, the CSS ground-truth database (Grant et al., 1993) can be of great use in this area. This database is a collection of regional events that have been reviewed, and some of the facts about the events (event type, location, depth, and origin time) are known with a high degree of confidence.

ACKNOWLEDGMENTS

We would like to thank Howard Patton for his suggestions and comments in the preparation of this manuscript. We also thank Peter Murphy for his editorial comments.

REFERENCES

Altschuler, E., T. Williams, E. Ratner, F. Dowla, and F. Wooten (1994). Method of constrained global optimization. Physical Review Letters, Vol. 72, No. 17, pp. 2671-2674.

Cansi, Y. (1995). An automatic seismic event processing for detection and location: the PMCC method. Geophysical Research Letters, Vol. 22, No. 9, pp. 1021-1024.

Dowla, F. and L. Rogers (1995). Solving Problems in Environmental Engineering and Geosciences with Artificial Neural Networks, in press.

Duckworth, G. L. (1987). A robust algorithm for tracking of drifting acoustic arrays in the Arctic. Proceedings of the 21st Asilomar Conference on Signals, Systems, & Computers, IEEE Press, pp. 103-108.

Flinn, E. A. (1965). Confidence regions and error determinations for seismic event location. Reviews of Geophysics, Vol. 3, No. 1, pp. 157-185.

Foufoula-Georgio, E. and P. Kumar (Eds.) (1994). Wavelets in Geophysics, Academic Press, San Diego.

Grant, L., J. Coyne, and F. Ryall (1993). CSS Ground-Truth Database: Version 1 Handbook. Technical Report C93-05, SAIC, Arlington, VA.

Harris, D. B. and F. U. Dowla (1986). Seismic arrays for treaty verification. Energy and Technology Review, pp. 8-19.

Harris, D. B. (1991). A waveform correlation method for identifying quarry explosions. Bull. Seism. Soc. Am., Vol. 81, No. 6, pp. 2395-2418.

Herrmann, R. B. (1982). Digital processing of regional network data. Bull. Seism. Soc. Am., Vol. 72, No. 6, pp. S261-S276.

Hertz, H., Al Krough, and R. G. Palmer (1991). Introduction to the Theory of Neural Computation. Addison-Wesley Publishing Company, Redwood City, CA.

Israelsson, H. (1990). Correlation of waveforms from closely spaced regional events. Bull. Seism. Soc. Am., Vol. 80, No. 6, pp. 2177-2193.

Johansson, E., F. Dowla, and D. Goodman (1992). Backpropagation learning for multilayer feedforward neural networks using the conjugate gradient method. International Journal of Neural Systems, Vol. 2, No. 4, pp. 291-301.

Neidell, N. S. and M. T. Taner (1971). Semblance and other coherency measures for multichannel data. Geophysics, Vol. 36, No. 3, pp. 482-497.

Park, J., C. R. Lindberg, and F. L. Vernon (1987). Multitaper spectral analysis of high-frequency seismograms. Journal of Geophysical Research, Vol. 92, No. B12, pp. 12675-12684.

Pavlis, G. L. (1986). Appraising earthquake hypocenter location errors: a complete, practical approach for single-event locations. Bull. Seism. Soc. Am. Vol. 76, No. 6, pp. 1699-1717.

Riviere-Barbier, F. and L. T. Grant (1993). Identification and location of closely spaced mining events. Bull. Seism. Soc. Am., Vol. 83, No. 5, pp. 1527-1546.

Shearer, P. M. (1994). Global seismic event detection using a matched filter on long-period seismograms. Journal of Geophysical Research, Vol. 99, No. B7, pp. 13713-13725.

Thorbjarnardottir, B. S. and J. C. Pechmann (1987). Constraints on relative earthquake locations from cross-correlation of waveforms. Bull. Seism. Soc. Am., Vol. 77, No. 5, pp. 1626-1634.

Thurber, C. H. (1985). Nonlinear earthquake location: theory and examples. Bull. Seism. Soc. Am., Vol. 75, No. 3, pp. 779-790.

Veith, K. F. (1975). Refined hypocenters and accurate reliability estimates. Bull. Seism. Soc. Am., Vol. 65, No. 5, pp. 1199-1222.

Wirth, M. H., R. R. Blanford, and R. H. Shumway (1976). Automatic seismic array and network detection. Bull. Seism. Soc. Am., Vol. 66, No. 4, pp. 1375-1380.

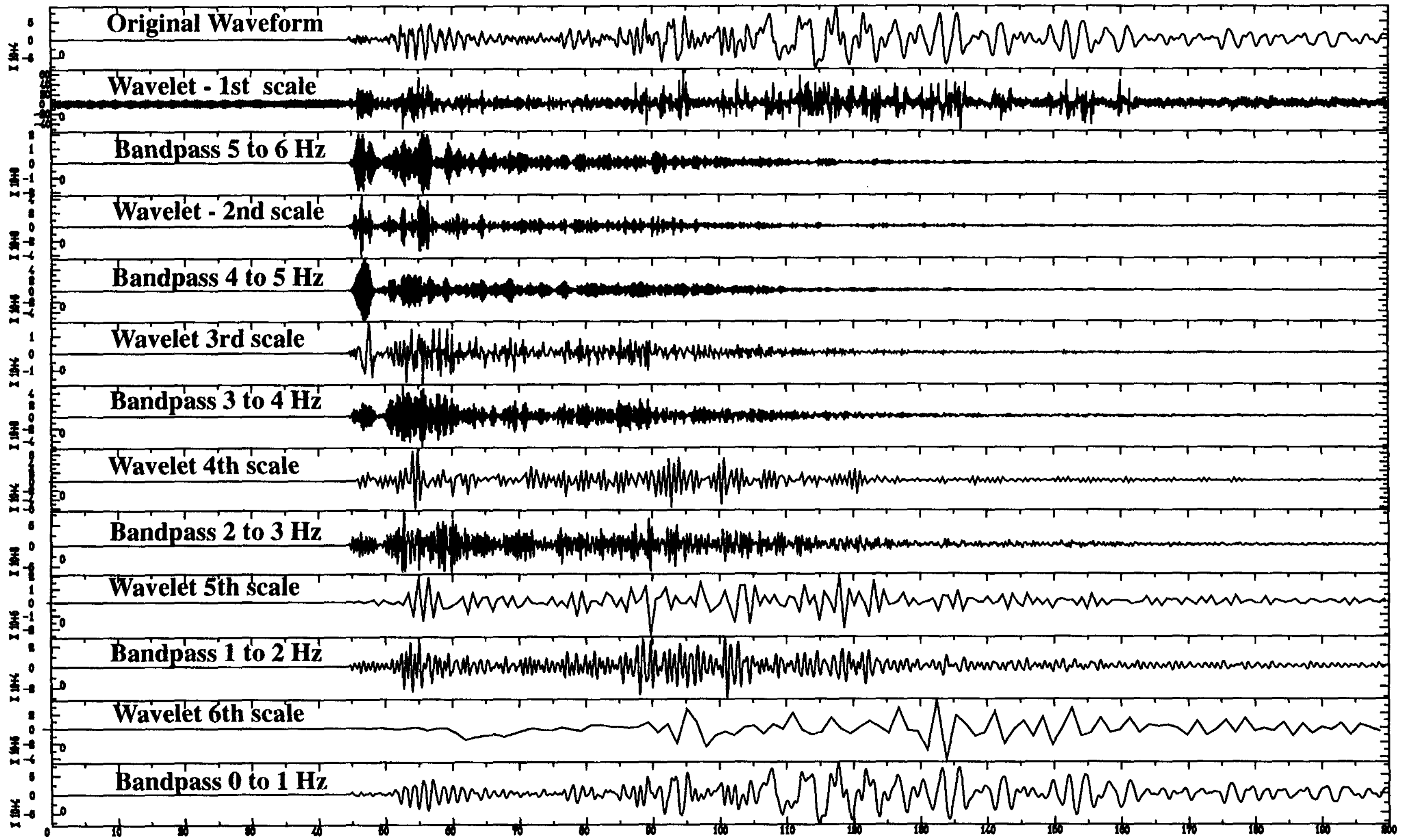


Figure 1. Comparison of wavelet coefficients and bandpass filtered seismogramsof an explosion record at Landers. The top signal is the orginal raw waveform. The bottom traces are alternate wavelet and bandpass representations at different scales and frequencies. Since the relation between scale and frequency is nonlinear, it is difficult to make direct comparison of the two techniques.

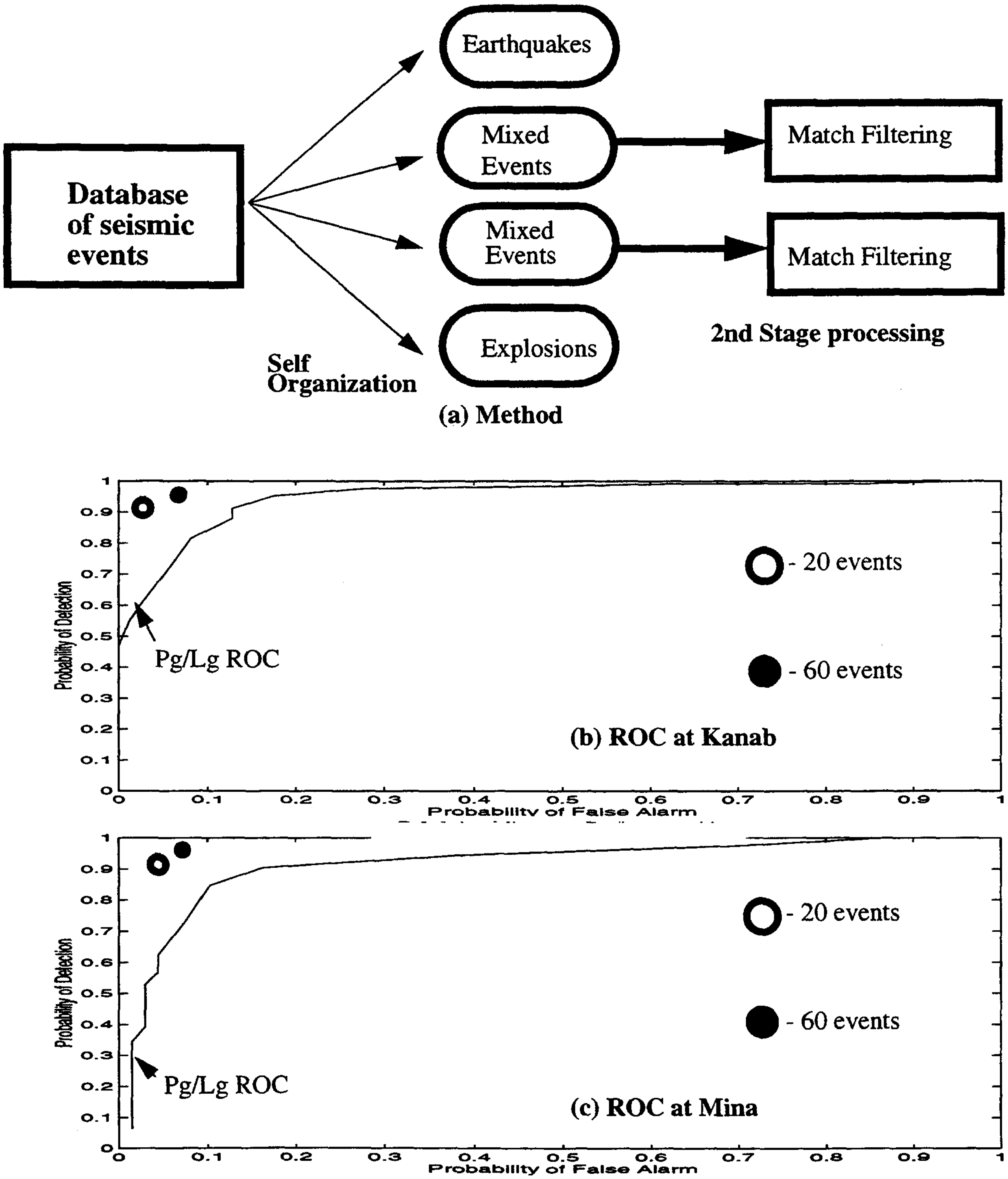
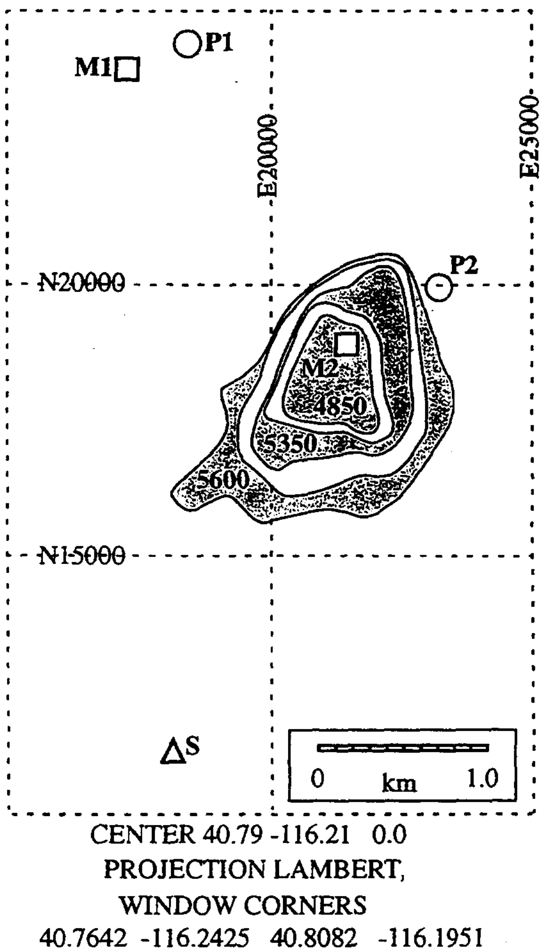
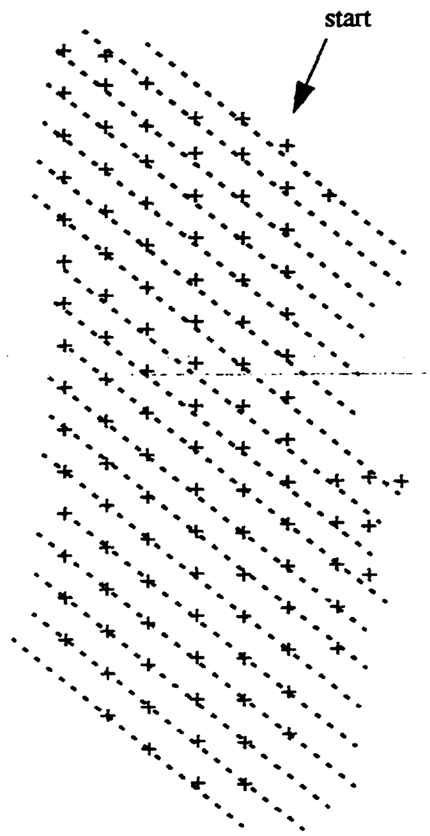


Figure 2. (a) The event discrimination algorithm described in the text is useful when there are many more earthquakes than explosions in the training database. (b) Performance comparison between the receiver operating characteristics (ROC) of a Pg/Lg discriminant and the self-organizing neural network with different number of explosions in the database at Kanab. (c) Same as (b) for the station Mina.



(a)



Firing pattern example

(b)

Figure 3. Map of the mining events recorded at the NGC gold quarry (left) showing the location of the April 14, 1995, explosion (M1) and the April 20, 1995, explosion (M2). Acoustic signals were recorded at P1 and P2 at a distance of about 0.3 km for events M1 and M2, respectively. Seismic signals were recorded at S for M1 and M2 at distances of 2.64 and 3.96 km. The shaded regions are the working levels in the Gold Quarry pit, which are at 4850-, 5350-, and 5600-ft elevations (above sea level). An example (not M1 or M2) firing pattern (right) shows the rows (dotted lines) and the individual explosion sites (+s). Delays between row detonations were approximately 50 ms for M1 and 65 ms for M2. Hole spacing for the individual explosions were 16x16 ft for M1 and 18x18 ft for M2.

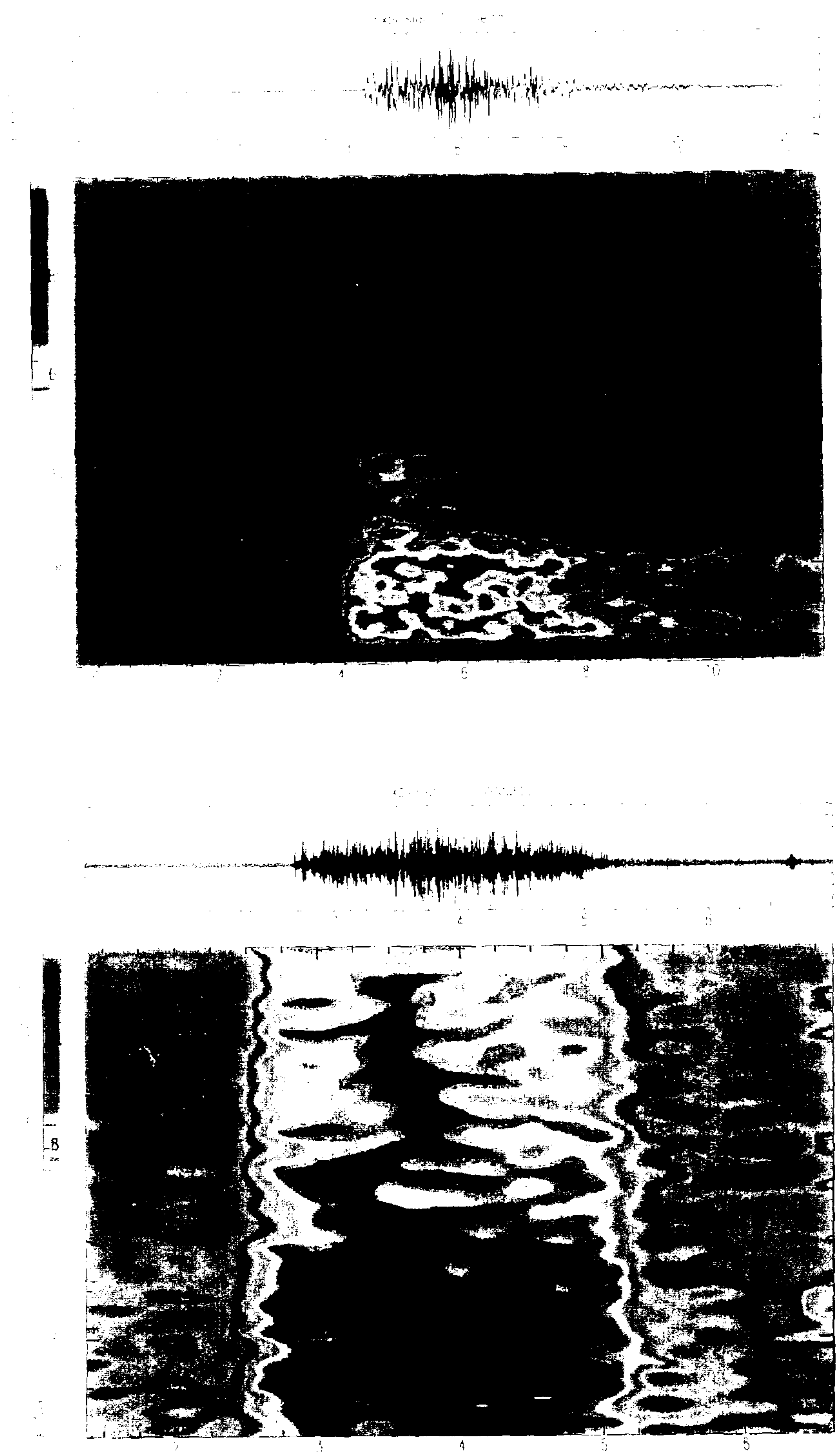


Figure 4. M1 mining explosion event showing filtered time series and resulting spectrogram for seismic signal (a) and for acoustic signal (b). All time series signals were bandpass filtered with a fourth order Butterworth filter. The seismic signal was bandpass limited from 1 to 100 Hz. The acoustic signal was bandpass limited from 500 to 1000 Hz.

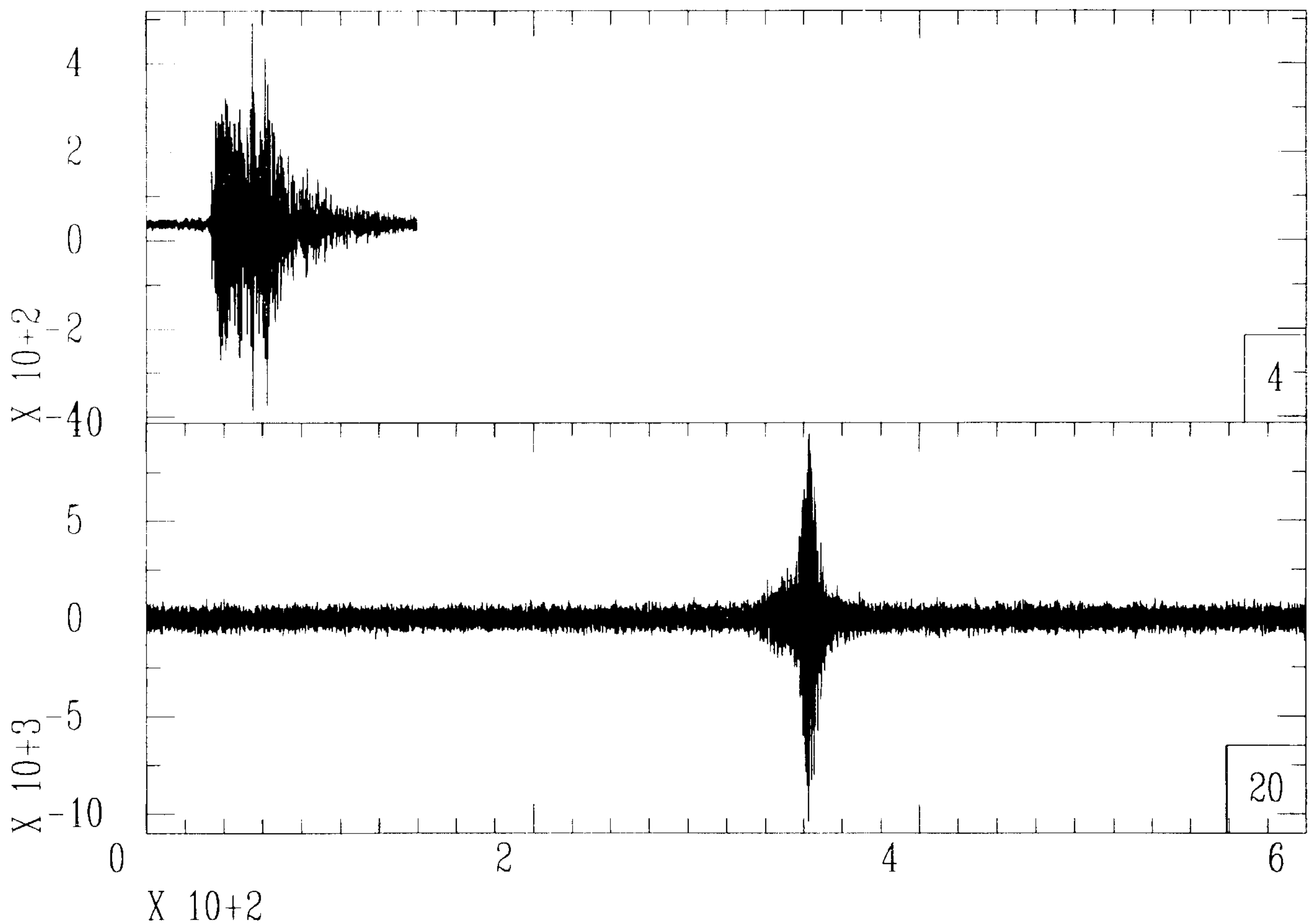


Figure 5. A qualitative comparison of seismic (top) and hydroacoustic (bottom) records for the same seismic event off the coast of southern California that occurred in 1994.